

The background of the slide is a dark, semi-transparent image of a Rocket League match. It shows a blue car on the left, a yellow car on the right, and a black car in the center, all in motion on a field. The text is overlaid on this image.

Predicting Rocket League Match Outcomes

Using Logistic Regression to Predict Match Results

Brandon Bailey
ITCS 3156 – Summer 2025

Project Overview

- What is Rocket League?
 - Rocket League is essentially soccer with flying cars, played in a fast-paced, competitive arena.
- What are we doing?
 - Using machine learning to predict the outcome of Rocket League Championship Series matches.
- Why?
 - Rocket League has simple rules but complex mechanics. This makes it perfect for modeling, because the environment remains the same. We can track the team behavior over time and use that to predict the winner
- How?
 - We extract match data (team stats, goals, assists) using tools like Ballchasing.com, then train a classification model and evaluate the prediction accuracy.

Motivation and Inspiration

Personal Interest

- RLCS 2025 – Raleigh Major | 6/28 - 6/29 | Prize Pool: \$351,000

Why Predict Rocket League Match Outcomes?

- Increase in Machine Learning in Esports

Ballchasing.com Replay Database

- Millions of data samples
- Various data formats

SaltieRL / carball

- Open-source replay decompiling
- Requirements
 - Python 3.6.7+ (3.7 and 3.8 included)

Related Works & Baseline Models

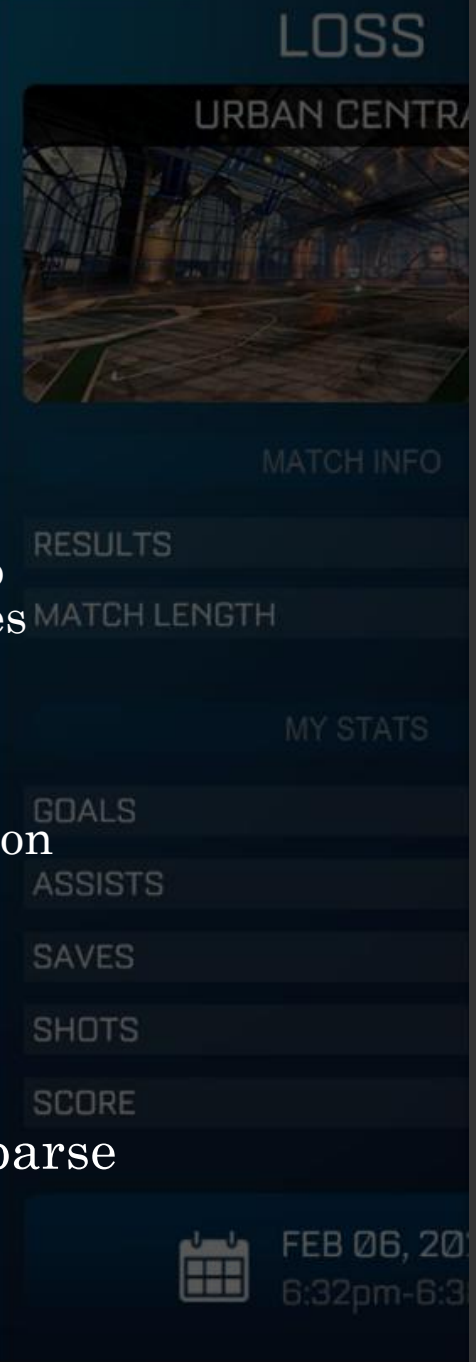
- Ballchasing.com and Carball Python package
 - Used in various projects to extract Rocket League replay data for ML
- [Predicting Rocket League Match Outcome With Machine Learning by: Walker Payne](#)
 - Walker's article shows how data scraping can give a competitive edge using ML
 - He built a model using 1v1 replays from Ballchasing.com and player ranking from Rocketleague.tracker
 - His goal was to predict match outcomes for use in Esports betting
 - This inspired me to build a similar model
 - Even though his model only had 57% accuracy, it was a proof of concept more than anything

Dataset Challenges

What Didn't Work

- Attempt to adapt [Walker's open-source 1v1 .replay scraper](#)
- When trying to extend the scraper to suit 3v3 RLCS data, it ran into formatting issues and was unable to successfully generate .XLSX files
- Carball Python package support
 - Unfortunately, Carball does not support Python 3.12, requiring Python 3.6.7 to 3.8.
 - Unable to downgrade python and continued to run into environment conflicts

Without Carball working, there is no way to efficiently extract/parse .replay files



More Dataset Challenges

What Also Didn't Work

- Kaggle Data Format Issues

- The original Kaggle data was split into two files:

- Frames.parquet: frame-by-frame player data (positions, boost, etc)

- Games.parquet: match-level data (teams, player names, game_id)

Problem: No reliable key to join the two files (**no game_id in frames file**)

- Merging frames.parquet and games.parquet:

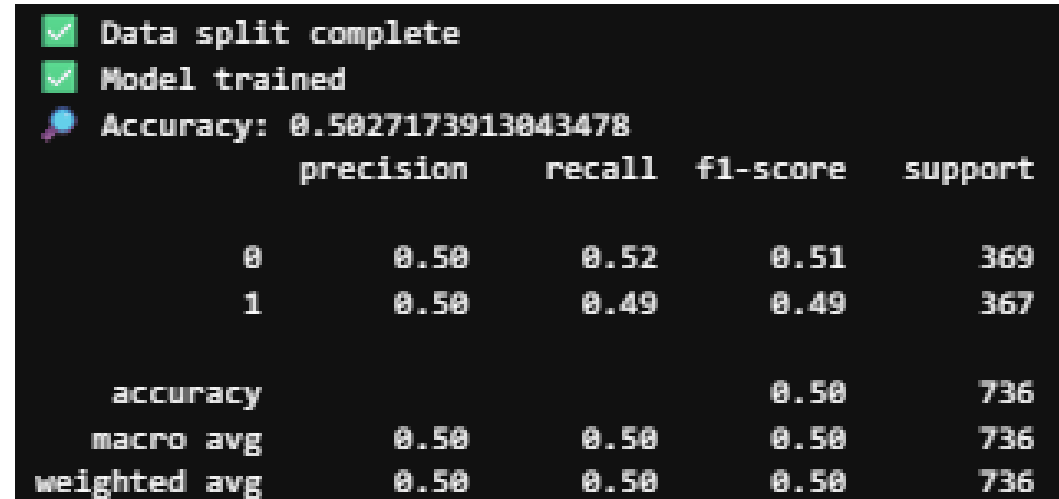
- Frames: (48957517, 105)

- Games: (17256, 15)

- Caused Memory Allocation Error

- Merging small portion of frames.parquet and games.parquet

- Trained a logistic regression model that performed poorly – 50% accuracy



```
✓ Data split complete
✓ Model trained
🚀 Accuracy: 0.5027173913043478
```

	precision	recall	f1-score	support
0	0.50	0.52	0.51	369
1	0.50	0.49	0.49	367
accuracy			0.50	736
macro avg	0.50	0.50	0.50	736
weighted avg	0.50	0.50	0.50	736

Working Dataset

Source: Ballchasing.com

- Search Filters
 - Pro: true (at least one pro player)
 - Playlist: ranked-standard (3v3)
- Each match includes a downloadable `-teams.csv` file with stats for both teams.
 - Downloaded 252 `-teams.csv` files
 - Each file has 2 rows (one per team)
- Usefulness:
 - No need to parse `.replay` files
 - Each file includes goals, shots, assists, saves, possessions, and many more features
 - Data is clean and consistent

Approach / Methods

Model Choice: Logistic Regression

- Chosen for its simplicity and interpretability
- Well-suited for binary classification (winner vs loser)

Other Model Options: Naïve Bayes

- Not chosen due to its assumption data is independent

Preprocessing Steps

Dropped Columns that could cause multicollinearity:

winner, replay_id

blue_color, orange_color

blue_team name, orange_team name

blue_score, orange_score

blue_goals, orange_goals

blue_goals conceded, orange_goals conceded

blue_shots conceded, orange_shots conceded

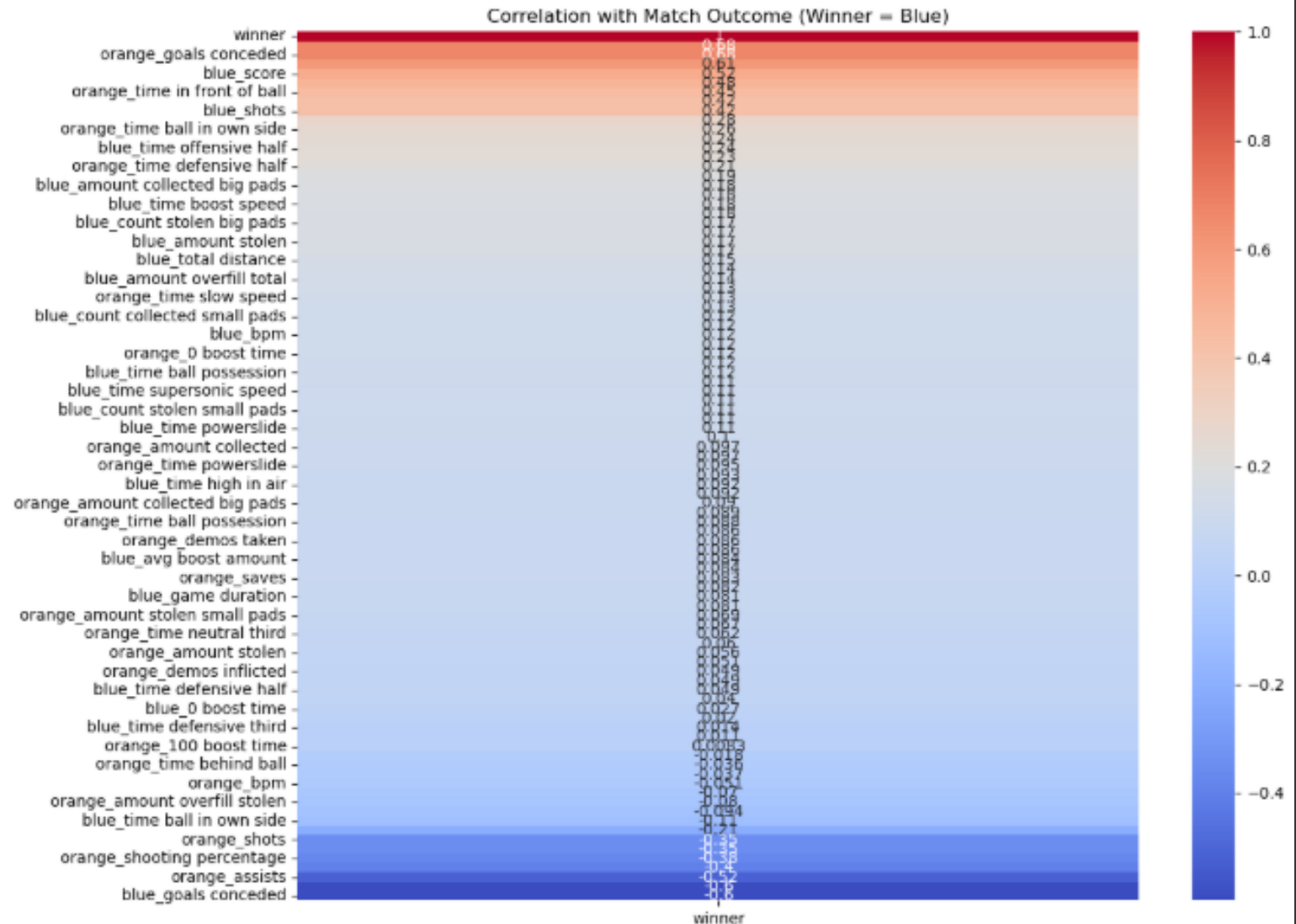
blue_assists, orange_assists

blue_saves, orange_saves

Experiments

Correlation with Match Outcome:

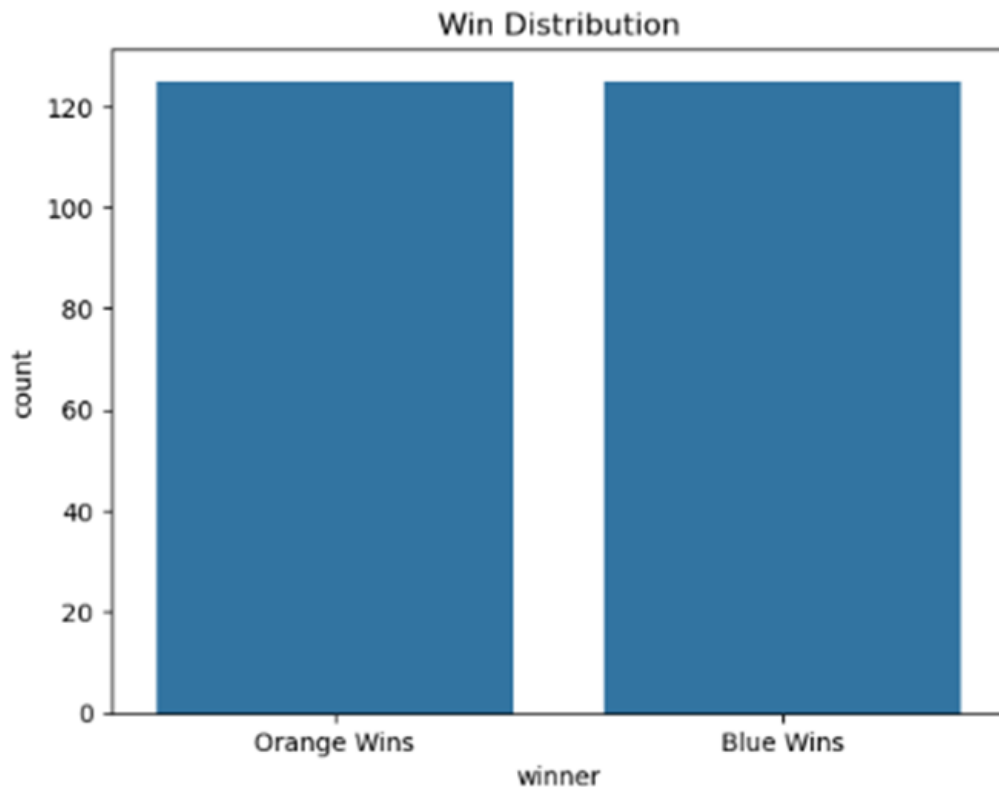
- Red = Positive correlation
(blue more likely to win)
- Blue = Negative correlation
(blue more likely to lose)



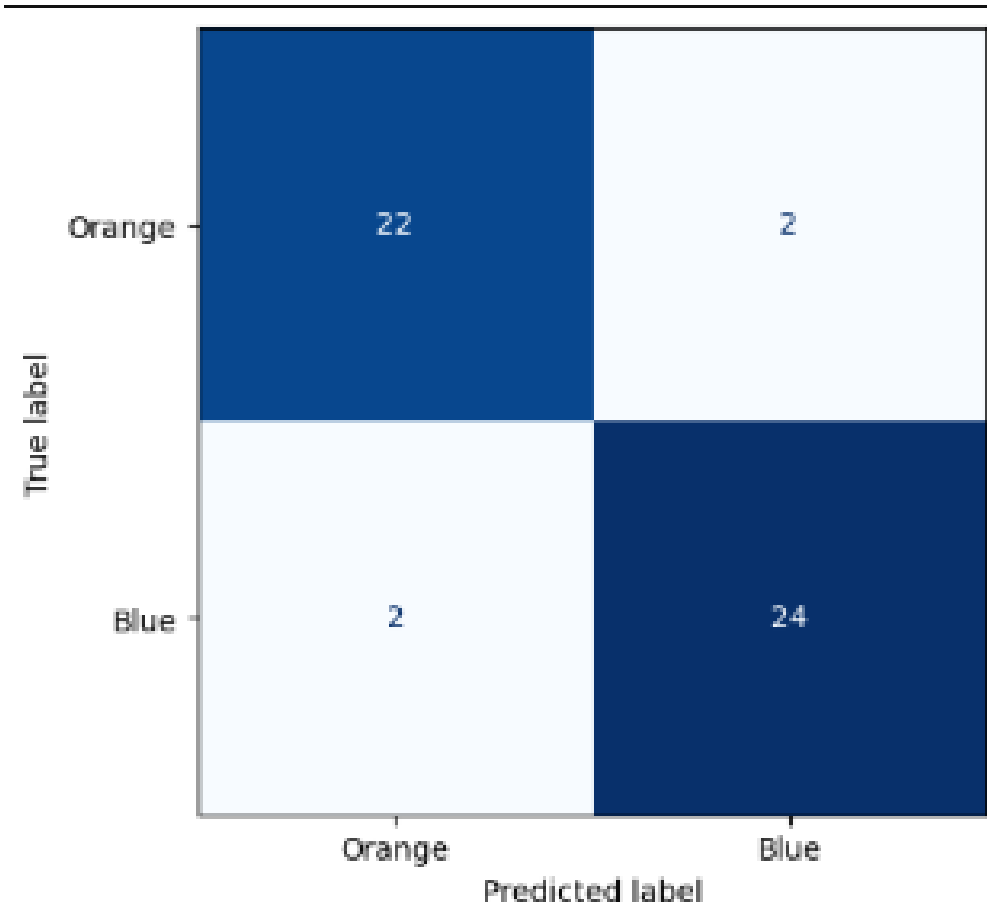
Experiments

Training Data Win Distribution

- Represents balance in training data



Confusion Matrix shows balanced prediction



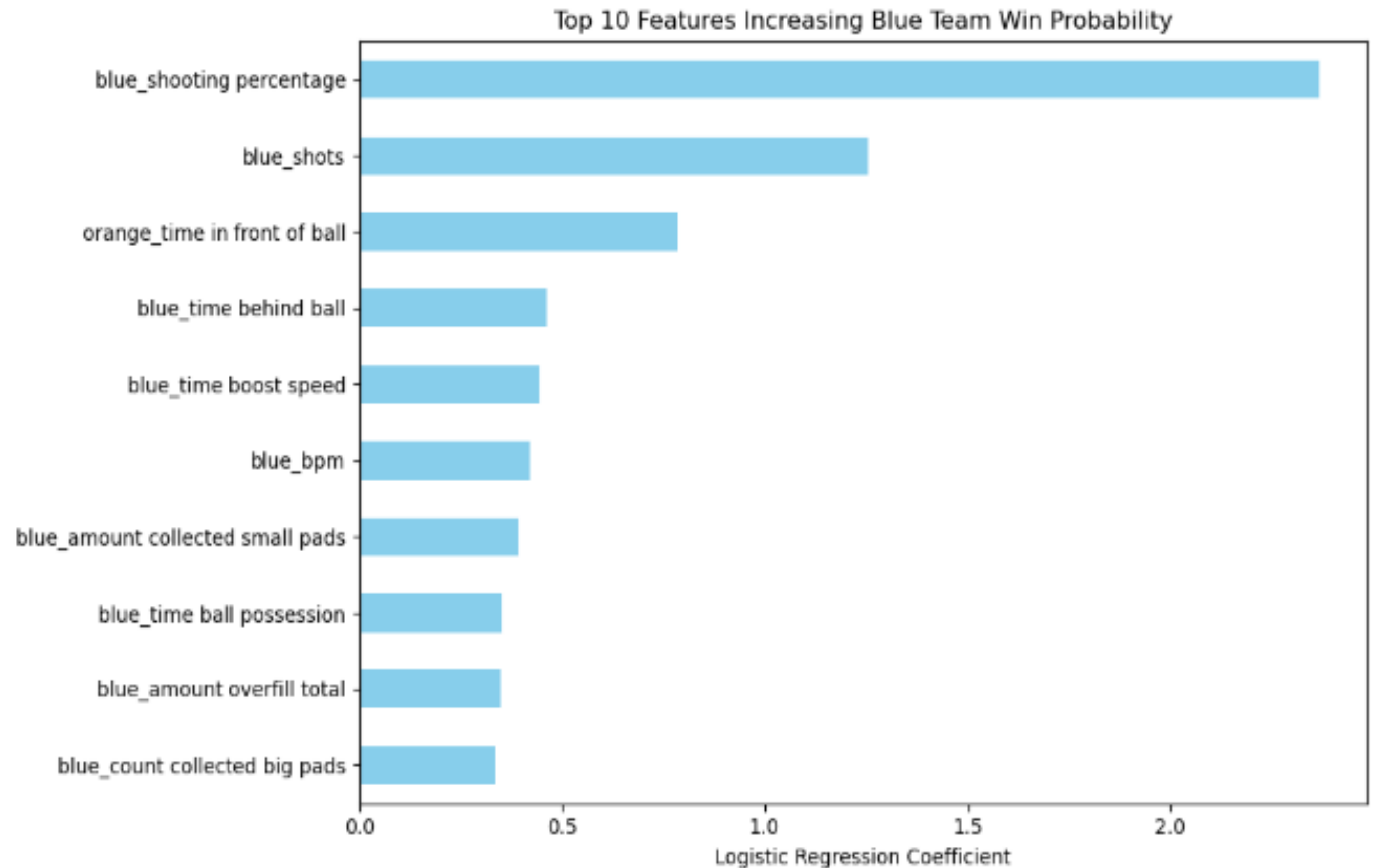
Key Results

- Final Model: Logistic Regression
- Test Accuracy: 92%
- Top Predictive Features
 - Blue shooting percentage
 - Blue shots
 - Orange time in front of ball

```
✓ Accuracy: 0.92
          precision    recall  f1-score   support

     0       0.92       0.92       0.92        24
     1       0.92       0.92       0.92        26

 accuracy          0.92          50
 macro avg         0.92       0.92       0.92        50
 weighted avg      0.92       0.92       0.92        50
```



Key Results

- Match Predictions

```
Analyzing Match ID: Raleigh Major - Team Falcons vs NRG Game 6 - 2025-06-29 -teams.csv
Predicted Winner: BLUE
Confidence - Blue: 0.68, Orange: 0.32
```

```
Top BLUE features:
orange_shooting percentage    3.184425
blue_time in front of ball    1.348346
blue_shots                    0.568478
dtype: float64
```

```
Top ORANGE features:
blue_shooting percentage    -2.373322
orange_shots                -2.211584
blue_amount overfill total  -0.363588
dtype: float64
```

```
Analyzing Match ID: Raleigh Major - Twisted Minds vs NRG Game 7 - 2025-06-29 -teams.csv
Predicted Winner: ORANGE
Confidence - Blue: 0.41, Orange: 0.59
```










```
Top BLUE features:
orange_shooting percentage    1.141548
blue_amount overfill total    1.869895
blue_count collected big pads 0.419812
dtype: float64
```

```
Top ORANGE features:
blue_shooting percentage    -1.284311
orange_shots                -1.111294
blue_shots                  -1.061566
dtype: float64
```










1 - TEAM FALCONS

		SCORE	GOALS	ASSISTS	SAVES	SHOTS
 Kiileerrz  	 Fennec	500	1	0	4	3
 Trk511 	 Fennec	309	0	0	2	4
 Rw9 	 Fennec	258	0	0	2	2











0 - NRG ESPORTS

		SCORE	GOALS	ASSISTS	SAVES	SHOTS
 Daniel 	 Fennec	296	0	0	2	6
 BeastMode 	 Octane	284	0	0	2	2
 Atomic 	 Octane	256	0	0	2	4

1 - TWISTED MINDS

		SCORE	GOALS	ASSISTS	SAVES	SHOTS
 rise. 	 Octane	455	1	0	3	2
 LTK_AtomiK 	 Fennec	377	0	0	3	0
 Nwpo. 	 Fennec	357	0	0	2	3

2 - NRG ESPORTS

		SCORE	GOALS	ASSISTS	SAVES	SHOTS
 Atomic  	 Octane	480	1	1	1	4
 BeastMode 	 Octane	353	1	0	1	2
 Daniel 	 Fennec	276	0	1	0	4

Future Work

Expand dataset

- Automate .replay scraping to include more matches
- Include frame series data for deeper context

Early Game Predictions

- Produce predictions given only half duration of a match

Hyperparameter Tuning

- Improve model generalization

Conclusion

Built a machine learning pipeline using real Rocket League match data

Learned:

- How to clean and structure esports data
- Why data and model compatibility is critical
- Train and evaluate a classification model using logistic regression

Success Criteria Met

- Model achieves greater than 70% accuracy on test set
- Model identifies reasonable features
- Results are visually clear

Failed Criteria

- Model can predict outcomes using only half a game's duration of data

Overall, this project proves that match outcome prediction is possible. With future improvements, there is a possibility for real-time predictions.

Sources

What is Rocket League:

- https://assets.nfhs.org/umbraco/media/1020429/what-is-rocket-league_.pdf

Walker Payne Medium Article

- [Predicting Rocket League Match Outcome With Machine Learning by: Walker Payne](#)
- Images
 - Title -
<https://image.api.playstation.com/vulcan/img/rnd/202009/2919/KpFO3I0iQCym2X58b43Avg8L.jpg>
 - Conclusion - https://wolfsgamingblog.com/wp-content/uploads/2015/07/2015-07-25_00005.jpg